ISSN 2464-4617(print) ISSN 2464-4625(CD) CSRN 2703 Computer Science Research Notes
http://www.WSCG.eu

Improved Adaptive Background Subtraction Method Using Pixel-based Segmenter

Kemal Batuhan Baskurt
Department of Computer Engineering
06830, Ankara, Turkey
batuhanbaskurt@gmail.com

Refik Samet
Department of Computer Engineering
06830, Ankara, Turkey
samet@eng.ankara.edu.tr

ABSTRACT

Moving object detection is essential in many computer vision systems as it is generally first process which feeds following algorithmic steps after getting camera stream. Thus quality of moving object detection is crucial for success of the whole process flow. It has been studied in the literature over the last two decades but it is still challenging issue because of factors such as background complexity, illumination variations, noise, occlusion and run-time performance requirement considering rapidly increasing image size and quality. In this paper, we try to contribute to solve this problem by improving an existing real-time non-parametric moving object detection method. In scope of this study, pixel based background model in which each pixel is represented separately by its distribution on time domain is used. Mentioned discrete background model is suitable for parallel processing by dividing the image to sub images in order to accelerate the process. Main feature of proposed non-parametric approach is automatic adjustment of algorithm parameters according to changes on the scene. This feature provides easy adaptation to environmental change and robustness for different scenes with unique parameter initialization. Another contribution is scene change detector to handle sudden illumination changes and adopt the background model to new scene in the fastest way. Experiments on 2012 ChangeDetection.net dataset show that our approach outperforms most state-of-the-art methods. Improvement obtained both on robustness and practical performance provides our approach to be able to use in real world monitoring systems.

Keywords

Real-time systems, computer vision, video surveillance, video processing, moving object detection, background subtraction

1. INTRODUCTION

Rapid incensement on using surveillance cameras in daily life has resulted in the need to find effective methods and algorithms to overcome huge data gathered every second. Moving object detection is one of the most commonly used methods to give the meaning of the raw data. A popular approach to solve moving object detection problem is background subtraction which has been studied in the literature over the last two decades. The idea of background subtraction is calculating difference between current frame and the background model which represents the scene regarding to data obtained from previous frames. A complete background subtraction method has three main components. First one is background modelling which tries to represent the scene characteristic in best way. Second is subtraction operation which indicates the method to calculate difference between background model and current frame. Third one is background update mechanism that provides adaptation to scene changes.

Background subtraction is a challenging problem since background might include large image variations due to lightening, repetitive motions, crowded scene and occlusions. These environmental difficulties make the background modelling complex and time-consuming. Besides handling problems mentioned above, run-time performance of background subtraction method is also important considering high quality images gathered from surveillance cameras.

First approaches on background subtraction in literature focused on static background model. The model contains just one image of the scene. Each pixel of received image compared with related pixel on background image to determine if it belongs to the background. While this approach might useful for analyzing short video sequences in controlled environment, it cannot handle multiple backgrounds like waving flags or trees. Therefore researchers worked on more sophisticated statistical background models such as Gaussian Mixture Model [Sta99], codebook [Kim05] or [Kae02]. Other authors have worked on other approaches like using collected pixel values instead of generation statistical model [Wan07, Bar11, Van12 and Hof12]. While some approaches use low-level features such as color and texture [Zha06, Kri06 and Jia08], sub-pixel edge map [Jai07], Sobel edges [Aza10], others try to solve problem using high-level semantic information of the scene on convolutional neural networks [Bra16]. There are wide scale surveys discussing theoretical backgrounds and evaluating run-time performance of background subtraction methods [Goy12, Sob14 and Vac12].

Our approach is based on PBAS method [Hof12], the differences lie in the neighbor update mechanism, automatic adjustment of increment/decrement of scene adaptation parameters and scene change detector algorithm for sudden illumination variations. Run-time performance of our approach is also improved by dividing the operation on sub-images under favor of discrete background model.

The paper is organized as follows. Section II describes related background subtraction methods. In section III, details of our approach are presented. Experiments and discussions are provided in section IV and section V concludes the paper.

2. RELATED WORK

Background subtraction methods aim to subtract moving object from static background without any priori information. Many methods have been proposed and extended survey papers can be found on this topic [Goy12, Sob14 and Vac12]. Existing methods can be divided into two groups as simple frame based methods and pixel based modelling methods.

Frame based methods also can be mentioned frame difference methods which use single image as background model. There are different approaches on building background image. Some approaches uses an image captured when there is no motion on the scene. Others simply calculate the difference between consecutive frames which means previous frame is always used as background model. [Lai98] describes background image by arithmetic mean of frames gathered at the training stage of their method. Absolute difference of background image and current image is used to determine motion area of the scene on all mentioned frame based approaches. They are unimodal approaches that background of each pixel is modelled by single value. These approaches are fast and easy to operate and efficient to detect instant motion on low-dynamic scenes. However frame based methods cannot handle dynamic background and long-term changes on the scene. Therefore more complex background models are proposed to solve environmental problems.

Over the years, several complex pixel level algorithms have been proposed. Most popular is Gaussian Mixture Model (GMM) [Sta99] which consists of modelling the distribution of the values observed over time at each pixel by a weighted mixture of Gaussians. This model handles most of the problems occurred because of the lack of multimodal background representation. Since its introduction, this model has been based by a lot of researchers and improved methods have been proposed. Main problem of GMM is high computational cost that prevents effective real-time operation of the method.

Another popular method is Codebook [Kim05] which models each pixel by a codebook which is a compressed form of background model. Each codebook contains codewords comprising colors transformed by an innovative color distortion metric. The method creates codebook model on training phase, then each frame is compared with the codebook model on test phase. Training phase of the method avoids adaptation to dynamic scene as significant changes on the scene after training cannot be handled. Necessity of training phase also makes the method inefficient in the meaning of easy-to-use. SACON [Wan07] method brought a new 'non-parametric' perspective by using collection of most recent image values at each pixel instead of statistical approach on background modelling. Each pixel of current frame is compared with stored collection of its previous values. Current pixel is labelled as background when it is 90% consistent with the background model. Oldest component of the background model is updated with new pixel value on update mechanism. ViBE [Bar11] and ViBE+ [Van12] use same background model with SACON but random component of the model is updated instead of the oldest one. Their decision criteria for background labelling are just 2 match with the background collection that makes the method faster than others. Mentioned fast update mechanism is built on conservative principle in which background model is only updated by background pixels.

Another non-parametric method PBAS [Hof12] also uses similar background model with ViBE but the randomness and decision thresholds are not fixed for all pixels as ViBE. Algorithm parameters are set separately for each pixel and they are changed dynamically according to scene variations over time. Mentioned dynamic parameter infrastructure makes the method more consistent regarding to scene changes while run-time performance decreases because of extra computational cost.

Meanwhile [Bra16] carries background subtraction on a different domain to solve the problem with spatial features learned with convolutional neural networks. Background model is generated by a single image and scene-specific training dataset. Their study indicates potential of deep features learned with conventional neural network for background subtraction without intention of proposing real-time and adaptive technique.

Our approach tries to improve deficiencies of PBAS method. Automatic adjustment of scene adaptation increment and decrement parameters which are used fixed in PBAS is added. Sudden illumination changes such as cloud passes, explosions caused by headlight or lightening variations on day-night change are important problems on real world applications. Background model is distorted by these artefacts and handling the distortion takes time with normal update mechanism of the method. Our approach contains scene change detector algorithm to cope with this problem. Background model is updated in fastest way once sudden illumination change detected. Scene change detector provides stability of our approach against uncontrolled environmental changes on long term analysis. Neighbor update mechanism of PBAS is also changed in our approach to avoid possibility of adding foreground pixels to background model. Finally image is divided into sub-images and the algorithm is applied in parallel through discrete structure of background model considering sub-image borders on neighbor update mechanism.

3. PIXEL-BASED ADAPTIVE SEGMENTER

Background subtraction method used in our approach consists of five main steps as follows:

a) Background/foreground decision
b) Background update
c) Dynamic update of decision threshold
d) Dynamic update of background update rate
e) Scene change detector

Proposed approach uses background model proposed by [Hof12]. Our approach differs from the base model on update mechanism, parameter dynamism and completely new scene change detector.

First of all, initial background model is created in our approach. Each pixel of new video frame is compared with its background model in order to decide whether it is background or not. Then background model is updated if the current pixel is labelled as background. Decision threshold used in first step and background update frequency are dynamically adjusted according to scene variation. Finally scene change detector is used to handle sudden dominant changes that distorts reliability of background model. Technical details and contribution of our approach on each step are explained in this section.

3.1 Background/Foreground Decision

Background/foreground decision step aims to compare each pixel with its background model and decide whether it is background or foreground pixel. Background model of a pixel $B(x_i)$ represents N recently observed pixel values:

$$B(x_i) = \{B_1(x_i), B_2(x_i), \dots, B_N(x_i)\} \quad (1)$$

A pixel is labelled as background when it's current value ($I(x_i)$) is closer than decision threshold ($R(x_i)$) at least minimum number ($\#_{min}$) of the N background samples, as shown in Fig. 1. Thus decision threshold represents distance between current pixel's value and background samples in colour space of input image.

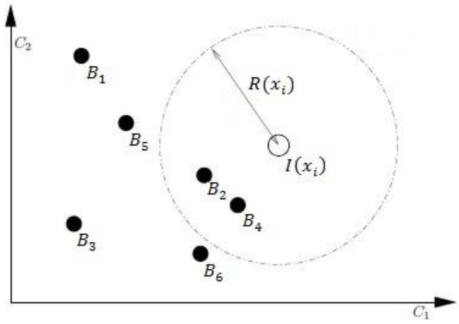


Figure 1. Background/foreground decision for 2-dimension (C_1, C_2) colour space

3.2 Background Update

Background model of a pixel is updated if it is labelled as background. Update operation is carried out by assigning current pixel value $I(x_i)$ to random selected sample $B_k(x_i)$ ($k \in 1, 2, \dots, N$). Current situation of the scene is learnt by background model in this way. Learning operation must be performed according to the scene change frequency. Thus, background model is updated in $p = 1/T(x_i)$ frequency instead of each frame. $T(x_i)$ represents pixel-based update rate which is adjusted dynamically with regarding to scene variation (see Section 3.4 for $T(x_i)$ definition).

Background model of random selected 8-connected neighbour of updated pixel ($y_i \in Neig(x_i)$) is also updated by neighbour's current pixel value in [Hof12]. Our approach proposes updating random selected neighbour's background with pixel value which is labelled as background (Eq. 2).

$$B_k(y_i) \leftarrow I(x_i) \quad (2)$$

Updating background model with the neighbour's pixel value is not appropriate as it may be labelled as foreground. Updating background model with foreground pixel value is prevented in our approach as it is inconsistent with principle of conservative background model.

3.3 Dynamic Update of Decision Threshold

Scene may have dynamic and stable regions at once, thus using fixed decision threshold and update rate for all pixels is insufficient considering real world scenarios. Decision threshold must be higher for dynamic regions that mean possibility of labelling moving pixels as background must be low. On the other hand, smaller changes on stable region must be considered for foreground with low decision threshold.

Minimum distance vector $D(x_i)$ between each updated background sample ($B(x_i)$) and current pixel value is stored to calculate pixel dynamism which provides to adjust decision threshold according to pixel changes on time domain.

$$D(x_i) = \{D_1(x_i), \dots, D_k(x_i), \dots, D_N(x_i)\} \quad (3)$$

$$d_{min}(x_i) = min_k dist\big(I(x_k), B_k(x_i)\big) \quad (4)$$

$$D_k(x_i) \leftarrow d_{min}(x_i) \quad (5)$$

Pixel dynamism is represented by average of minimum distance values for all background samples of the pixel. $R(x_i)$ is increased/decreased by increment/decrement parameter ($R_{inc/dec}$) when dynamism reaches to upper limit which is determined by R_{scale} parameter in Eq. 6.

$$R(x_i) = \begin{cases} R(x_i) \cdot (1 - R_{inc/dec}), if\ R(x_i) > \bar{d}_{min}(x_i) \cdot R_{scale} \\ R(x_i) \cdot (1 + R_{inc/dec}), \quad else \end{cases} \quad (6)$$

$R(x_i)$ is limited by lower decision value (R_{lower}) in order to control decision criteria in acceptable limits.

3.4 Dynamic Update of Background Update Rate

Background model update rate $T(x_i)$ which represents update frequency (in frames) of the model is another pixel-based adaptive parameter related to pixel dynamism. Background model of high-dynamic region is updated rare than stable region for preserving background model from moving objects. Update rate of foreground pixel is increased for rare update on the region. $1/\,T(x_i)$ is update frequency where $T(x_i)$ is calculated as follows:

$$T(x_i) = \begin{cases} T(x_i) + \dfrac{T_{inc}}{\bar{d}_{min}(x_i)}, & if\ F(x_i) = 1 \\ T(x_i) - \dfrac{T_{dec}}{\bar{d}_{min}(x_i)}, & if\ F(x_i) = 0 \end{cases} \quad (7)$$

where $(F(x_i) = 1)$ represents foreground pixel, while $(F(x_i) = 0)$ represents background. $T(x_i)$ parameter is limited between minimum (T_{lower}) and maximum (T_{upper}) values in order to control the background model in the case of false updating.

Increment/decrement parameters of both decision threshold ($R_{inc/dec}$) and update rate (T_{inc}, T_{dec}) are also adjusted dynamically in our approach while they are fixed in [Hof12]. Using fixed increment/decrement step for all pixels during entire run-time causes slow reaction of the method over fast changes in the scene. Our approach on this point accomplishes complete adaptation to dynamic scene. Each pixel uses its own increment/decrement parameter instead of unique ones for all. Increment/decrement parameters are changed in 1 % ratio according to dynamism of the pixel for each step. Mentioned parameters are increased for stable pixels while decreased for dynamic pixels. In the case of sudden change on stable region, threshold parameters are quickly adopted to new scene because of bigger increment/decrement steps.

$$R_{inc/dec}(x_i) = \begin{cases} R_{inc/dec}(x_i) \cdot 0.99, if\ R(x_i) > \bar{d}_{min}(x_i) \cdot R_{scale} \\ R_{inc/dec}(x_i) \cdot 1.01, \quad else \end{cases} \quad (8)$$

$$T_{inc}(x_i) = \begin{cases} T_{inc}(x_i). 0.99, & if\ F(x_i) = 1 \\ T_{inc}(x_i). 1.01, & if\ F(x_i) = 0 \end{cases} \quad (9)$$

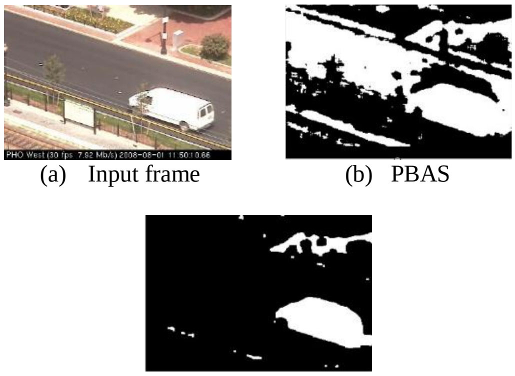
$$T_{dec}(x_i) = \begin{cases} T_{dec}(x_i). 0.99, & if\ F(x_i) = 1 \\ T_{dec}(x_i). 1.01, & if\ F(x_i) = 0 \end{cases} \quad (10)$$

3.5 Scene Change Detector

Dominant changes on the scene such as streetlight (de)activation on day/night change, cloud passing or photocell lightening cause serious problem for background subtraction methods even dynamic update rate is used. Normal adaptation process of background model to new scene structure takes long time which means a large number of false detection during this period. Our approach has a precaution named scene change detector for handling this unusual situation. Maximum motion ratio M_{Max} is defined to control unexpected dominant changes.
Update rate of foreground pixels is assigned to T_{lower} which means highest update frequency when the ratio of foreground pixels reaches to M_{Max}. Update ratio resumes to normal frequency, after then background model learns the scene in the fastest way.

$$T(x_i) = \begin{cases} T_{lower}, & \#_{(F=1)} > (M_{Max} \cdot \#_x) \\ T(x_i), & else \end{cases}, i \in (F(x_i) = 1) \quad (10)$$

Effect of scene change detector compared to PBAS result is showed in Fig 2.



(a) Input frame (b) PBAS

(c) Our Approach

Figure 2. Effect of scene change detector (818th frame of boulevard scenario on camera jitter category of change detection dataset)

Our approach is more adaptive than based approach proposed by [Hof12] considering dynamic adjustment of more parameters. Fewer parameters adjusted by user makes our approach more robust against different scenes. Scene change detector also solves common problem faced in real world applications.

4. IMPLEMENTATION AND EXPERIMENTAL RESULTS

Following metrics are used to evaluate performance of proposed approach (Table 1).

Metric	Explanation
Recall	TP/ (TP + FN)
F1	(2 * Precision * Recall) / (Precision + Recall)
Precision	TP / (TP + FP)

Table 1. Performance Metrics (TP: True Positive, FP: False Positive, FN: False Negative, TN: True Negative)

Recall represents fraction of number of foreground pixels classified as foreground over number of foreground pixels classified as background. Precision represents fraction of number of foreground pixels classified as foreground over number of background pixels classified as foreground. F1 represents harmonic average of recall and precision.

Our approach is compared to PBAS (implementation provided by the authors) in terms of three performance measures (Table 2) on six scenarios provided by change detection 2012 benchmark [Goy12].

Run-time performance comparison of both methods is also provided in Table 2 (bold values are the best in the comparison). Change detection 2012 benchmark provides extensive comparison of 44 state-of-the-art methods including PBAS. Thus comparing our approach with PBAS on this dataset also provides opportunity to evaluate our performance among other state-of-the-art methods.

Optimal parameter setting proposed by the authors of PBAS is follows: {N = 35, $\#_{min}$ = 2, $R_{inc/dec}$ = 0.05, R_{lower} = 18, R_{scale} = 5, T_{inc} = 1, T_{dec} = 0.05, T_{lower} = 2, T_{upper} = 200}. Same parameter setting except T_{upper} = 5 is used for our approach. As mentioned in previous section $R_{inc/dec}$, T_{inc} and T_{dec} parameters are adjusted dynamically according to scene variation in our method. Thus defined values above are initial ones for these parameters. They are changed in the ratio of 0.01 according to dynamism of related pixel. Limitation parameters for these ones are used as follows: {$R_{inc/dec}^{upper}$ = 0.05, $R_{inc/dec}^{lower}$ = 0.01, T_{inc}^{upper} = 1.5, T_{inc}^{lower} = 0.5, T_{dec}^{upper} = 0.1, T_{dec}^{lower} = 0.02}

Result of each scenario with the overall of each performance measure is provided. Our approach shows better performance in all measures of overall.

As mentioned in Section III, discrete background model in which each pixel is represented separately by its recently obtained values is used in our approach. Discrete model provides opportunity of parallel processing by dividing scene to sub images and operate them separately. Important point to pay attention on parallelization is updating background model of neighbor pixel on intersection region of sub-images. Border control is added to avoid manipulation of others memory between threads. Our implementation divides the scene 2x2 sub-images for parallel processing. Run-time performance presented in Table 2 shows effectiveness of our implementation. Our approach is processed 61 % faster than PBAS on overall.

5. CONCLUSION

We have presented improvement of our approach across PBAS. Three more parameters are adjusted dynamically according to scene change instead of using constant values. Neighbour update mechanism is changed to prevent updating background model with foreground pixel values. Moreover our scene change detector algorithm provides fast adaptation to major changes on the scene without large number of false detection. Our approach also benefits from discrete structure of background model in order to parallelise the method on sub-images. Mentioned improvements both on algorithm and implementation outperform PBAS and most of state-of-the-art methods. Future work will focus on completely dynamic method without necessity of any constant parameter. Performance on intermittent object detection scenario which dramatically decreases the overall performance also seems to need improvement

		Baseline	Camera Jitter	Dynamic Background	Intermittent Object Motion	Shadow	Thermal	Overall
Recall	PBAS	**0,8259**	**0,7927**	**0,7918**	0,4409	0,8447	0,6395	0,7226
	Our Approach	0,8246	0,7114	0,7836	**0,5238**	**0,8665**	**0,6617**	**0,7286**
F1	PBAS	0,7820	0,5490	0,6183	0,4922	**0,7772**	0,6806	0,6499
	Our Approach	**0,8198**	**0,5617**	**0,6974**	**0,5379**	0,7550	**0,7176**	**0,6816**
Precision	PBAS	0,7698	0,4386	0,6535	**0,7206**	**0,7283**	0,7884	0,6832
	Our Approach	**0,8225**	**0,5211**	**0,7182**	0,6308	0,7130	**0,8176**	**0,7039**
Run Time (ms)	PBAS	25,25	33,50	32,00	16,50	23,67	17,00	24,65
	Our Approach	**19,75**	**18,5**	**17,33**	**10,5**	**15**	**10**	**15,18**

Table 2. Results of PBAS on all Scenarios of Change Detection Dataset

6. REFERENCES

[Aza10] Azab, M. M., Shedeed, H. A., & Hussein, A. S. "A new technique for background modeling and subtraction for motion detection in real-time videos, in Image Processing (ICIP), 2010 17th IEEE International Conference on (pp. 3453-3456). IEEE.

[Bar11] Barnich, O., & Van Droogenbroeck, M. "ViBe: A universal background subtraction algorithm for video sequences," Image Processing, IEEE Transactions on, 20(6), 1709-1724, 2011.

[Bra16] Braham, M., & Van Droogenbroeck, M. "Deep Background Subtraction with Scene-Specific Convolutional Neural Networks," in International Conference on Systems, Signals and Image Processing, Bratislava 2016. IEEE.

[Goy12] Goyette, N., Jodoin, P. M., Porikli, F., Konrad, J., & Ishwar, P. "Changedetection. net: A new change detection benchmark dataset," in Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on (pp. 1-8). IEEE.

[Hof12] Hofmann, M., Tiefenbacher, P., & Rigoll, G. "Background segmentation with feedback: The pixel-based adaptive segmenter," in Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on (pp. 38-43).

[Jai07] Jain, V., Kimia, B. B., & Mundy, J. L. "Background modeling based on subpixel edges," in Image Processing, 2007. ICIP 2007. IEEE International Conference on (Vol. 6, pp. VI-321). IEEE.

[Jia08] Jian, X., Xiao-qing, D., Sheng-jin, W., & You-shou, W. "Background subtraction based on a combination of texture, color and intensity," in Signal Processing, 2008. ICSP 2008. 9th International Conference on (pp. 1400-1405). IEEE.

[Kae02]KaewTraKulPong, P., & Bowden, R. "An improved adaptive background mixture model for real-time tracking with shadow detection," in Video-based surveillance systems (pp. 135-144). Springer US, 2002.

[Kim05] Kim, K., Chalidabhongse, T. H., Harwood, D., & Davis, L. "Real-time foreground–background segmentation using codebook model," Real-time imaging, 11(3), 172-185, 2005.

[Kri06] Kristensen, F., Nilsson, P., & Öwall, V. "Background segmentation beyond RGB," in Computer Vision–ACCV 2006 (pp. 602-612). Springer Berlin Heidelberg.

[Lai98] Lai, A. H., & Yung, N. H. "A fast and accurate scoreboard algorithm for estimating stationary backgrounds in an image sequence," in Circuits and Systems, 1998. ISCAS'98. Proceedings of the 1998 IEEE International Symposium on (Vol. 4, pp. 241-244). IEEE.

[Sob14] Sobral, A., & Vacavant, A. "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos," Computer Vision and Image Understanding, 122, 4-21, 2014

[Sta99] Stauffer, C., Grimson, W. E. L. "Adaptive background mixture models for real-time tracking," in Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on. (Vol. 2). IEEE.

[Vac12] Vacavant, A., Chateau, T., Wilhelm, A., & Lequièvre, L. "A benchmark dataset for outdoor foreground/background extraction," in Computer Vision-ACCV 2012 Workshops (pp. 291-300). Springer Berlin Heidelberg.

[Van12] Van Droogenbroeck, M., & Paquot, O. "Background subtraction: Experiments and improvements for ViBe." in Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on (pp. 32-37).

[Wan07]Wang, H., & Suter, D. "A consensus-based method for tracking: Modelling background scenario and foreground appearance," Pattern Recognition, 40(3), 1091-1105, 2007.

[Zha06] Zhang, H., & Xu, D. "Fusing color and texture features for background model," in Fuzzy Systems and Knowledge Discovery: Third International Conference, FSKD 2006, Xi'an, China, September 24-28, 2006. Proceedings (pp. 887-893). Springer Berlin Heidelberg.